T-PEMD-88-9

United States General Accounting Office

GAO

Testimony

For Release
on Delivery
Expected at
10:00 a.m.
Wednesday,
Sept. 14, 1988

WEAPONS TESTING: QUALITY OF DOD OPERATIONAL TESTING AND REPORTING

Statement of
Eleanor Chelimsky, Director
Program Evaluation and
Methodology Division

Before the
Acquisition Policy Panel
Committee on Armed Services
House of Representatives

GA 1.5/2:

Mr. Chairman and Members of the Panel: It is a pleasure to be here today to present GAO's findings on operational testing and evaluation (OT&E) in the Department of Defense (DOD). For today's hearing, we have prepared an unclassified overview of our report. To keep the testimony unclassified, my presentation will need to stay at a rather general level. However, if there are questions that require a classified or potentially classified response, I'll be happy to address them in a closed session at the Panel's convenience.

In 1983, the Congress established the office of the Director of Operational Test and Evaluation (DOT&E) to effect several reforms concerning operational testing.[1] Prominent among the reform objectives were independent oversight and coordination of the military services' planning and execution of operational tests, independent evaluation of the results of operational tests, and objective reporting of those results to decisionmakers in the Department of Defense (DOD) and the Congress. A fundamental congressional concern was that weapons were not being tested thoroughly or realistically and that complete and accurate information was not being disseminated.

The Chairman, Seapower Subcommittee, House Armed Services

---

[1]In practice, the acronym DOT&E is used to denote both the director and the office under his direction. To avoid confusion, we refer to the director as the director, to the office as DOT&E, and we use the term OT&E to refer to operational test and evaluation.

best copy available

Committee, and three other Members of Congress, asked us to address two evaluation questions: 1) What is the methodological adequacy of OT&E under DOT&E oversight?, and 2) What is the quality of DOT&E dissemination of information to the Congress? In answering question 1, we also made an effort to determine the impact of DOT&E on the OT&E process.

SCOPE AND METHODOLOGY

To address the questions, we reviewed relevant documentation on the OT&E of six weapon systems, as well as congressional testimony, DOD regulations, and outside literature on the conduct and reporting of OT&E in general. We also interviewed DOD officials and outside experts in OT&E. All field work was conducted between September 1987 and March 1988.

Case Selection

To select the six weapon systems, we developed objective case-selection criteria. These criteria and their rationales are shown in table 1. The use of these criteria yielded 10 eligible candidates. The final six were selected on the combined basis of recency and number of common missions. The latter was important to facilitate greater comparability across systems. Final selections were: for the Army, Army Helicopter Improvement Program (AHIP) and Aquila Remotely Piloted Vehicle (RPV); for the Navy, Conventional

Tomahawk Land Attack Missile (TLAM/C) and DDG-51 Destroyer (Aegis Anti-Air Warfare system only); and for the Air Force, Imaging Infrared (IR) Maverick and Low-Altitude Navigation and Targeting Infrared System for Night (LANTIRN).

Table 1: GAO's Initial Case Selection Criteria and Rationales

Criterion	Description	Rationale
1.	B-LRIP report filed or scheduled for FY 1987	The Beyond-Low Rate Initial Production (B-LRIP) report is DOT&E's system specific reporting requirement under 10 USC 138, and therefore is necessary to fully address the second evaluation question. Also, for cases which met this criteria, DOD testing to justify production will be complete.
2.	Entry into B-LRIP after director swear-in	There was no permanent DOT&E director until April, 1985. Cases that entered B-LRIP after that date were included.
3.	Must be a major system	The Congress is primarily interested in major systems, those for which a Selected Acquisition Report is required (that is, those over $200 million in research and development or $1 billion in production).
4.	Must be a non-strategic system	Congressional requesters expressed primary interest in conventional tactical systems (as opposed to strategic nuclear systems).
5.	Must have tri-service representation	DOT&E oversees testing across all of DOD. Sampling cases from each service allows DOD-wide conclusions.
6.	Study to include six systems maximum	Time and staff available limit the number of OT&Es and reports to the Congress that GAO can adequately evaluate.

As is evident from the selection criteria, this review focuses on OT&E of major, conventional systems that reached the B-LRIP milestone by the end of fiscal year 1987. Therefore, the results are generalizable only to that universe of systems; they are not generalizable to the entire universe of OT&E being conducted under DOT&E oversight, including strategic systems or systems that have not yet reached the B-LRIP milestone. In addition, the results do not permit a direct assessment of change attributable to the legislative establishment of DOT&E. Such an assessment would have required time, resources, and data that were beyond the scope of the present study.

Assessment Framework

During a prior evaluation of the Joint Test and Evaluation program (JT&E), we developed a multiple case study method to assess the quality of the tests. We later refined the method in our evaluation of the Joint Live Fire Test program (JLF) and used it again for the present evaluation of DOT&E. First, a standardized assessment framework was developed to evaluate the cases. Next, information on each case was analyzed and coded in terms of the assessment framework. To ensure appropriate and consistent interpretation of the framework, all coding was continually monitored and validated across cases. Lastly, the information from each case was synthesized across cases to yield overall findings and conclusions.

Sources used to develop the framework included 1) DOD regulations on the conduct and reporting of OT&E (DOD Directives 5000.3 and 5000.3-M-1), 2) statements by the DOT&E director from congressional testimony, 3) the legislation which established DOT&E, 4) prior studies on OT&E, and 5) the JT&E and JLF assessment frameworks. The assessment framework covered seven categories: planning, execution, realism, analysis, reporting by the service operational test agencies, DOT&E impact, and DOT&E reporting. Each category contained a set of assessment questions or items. We stress that the items in our framework and their interpretation were based on established DOD guidance. For example, the importance of a realistic portrayal of threat forces is noted in Directive 5000.3, in DOT&E's own statements, and in prior studies on OT&E. And in each case, we compared the threat as portrayed in the OT&E to the threat as portrayed in DOD-approved threat assessments.

METHODOLOGICAL ADEQUACY OF OT&E UNDER DOT&E OVERSIGHT

Congressional concern regarding the inadequacy of operational testing of weapons under realistic, combat-like conditions was a principal reason for the establishment of DOT&E. In addressing methodological adequacy, we focused on whether significant problems or limitations were found in each OT&E. We define a significant problem or limitation as one that potentially affects conclusions

regarding the operational effectiveness or suitability of the weapon system. We do not report unimportant problems and limitations, those that in our judgment do not meet this criterion. Moreover, it is important to recognize that many problems and limitations in OT&E are unavoidable. Due to time, resource, and safety constraints, not everything can be tested or tested well. Further, it is not our intention to hold either DOT&E or the service test agencies responsible for events they cannot control.

To keep the testimony unclassified, system identifiers are not included (this section only).

Planning and Execution

Findings on planning and execution are summarized in table 2. DOD Directive 5000.3 (the principal regulation for DOD testing and evaluation) states that the test plan must provide a clear correlation between critical issues and test objectives through test-verifiable criteria. However, we found significant problems in five of the six cases involving the extent to which this was done. Criteria were either nonexistent, or stated qualitatively when they could have been stated quantitatively, or did not show clear, meaningful relationships with objectives. DOD Directive 5000.3 M-1 (supplemental TEMP guidelines) also states that test criteria must reflect the performance and limitations of other components that support that mission. Nevertheless, we found

criteria that did not consider the limitations of other systems that the system being tested must operate with. This resulted in some tests in which the tested system was successful but the mission it was designed to carry out was not.

Table 2:

Significant Problems and Limitations in Test Planning and Execution

	Army systems		Navy systems		Air Force systems	
Assessment questions	AHIP	Aquila	TLAM/C	DDG-51	IR Maverick	LANTIRN
Planning						
Did TEMP include a complete statement of the system's requirements?	a	X			a	
Did test plan address all system requirements and critical issues identified in the TEMP?	a		X	X	a	X
Was there a clear relationship in the test plan between critical issues and test objectives through meaningful, test verifiable criteria?		X	X	X	X	X
Execution						
Was each system requirement and critical issue identified in the test plan tested for as planned?	X	X	X			
Were there limitations in implementation that had not been anticipated in the test plan?	X	X	X	X		

Note: empty cells signify "no significant problems or limitations found." X signifies "one or more significant problems or limitations found." a signifies "insufficient information to evaluate."

In three of six cases, requirements and critical issues were not tested as specified in the test plan. In one of these, the plan was significantly changed after initial approval, resulting in a test in which the execution of the originally planned objectives and the added objectives was significantly diluted. Changes in the other cases removed significant aspects of the realism with which systems or crews were stressed. (These are discussed below under realism.) We also found significant problems and limitations that had not been anticipated in the test plan in four of six cases. In three of these, the result was a reduction in realism that favored the system being tested.

Realism

Findings on realism are summarized in table 3. Directive 5000.3 states that typical users should operate and maintain the system, and prior OT&E studies, as well as the law that established DOT&E, also stress the importance of typical users. We found that in four of six cases the system operators were not typical, and that in four of five cases where the question was applicable, the support personnel were not all typical. The prevailing problem for operational users was that they were selected from an operator pool that was atypically high in their skill or experience level--that is, so-called "golden crews." The prevailing problem for support personnel was some level of contractor involvement in the support of the system, principally in the maintenance function, although

that contractor support would not be available in the field. In addition to being organizationally different, contractor maintenance personnel are usually better trained and more experienced on the system than military personnel would be. Consequently, their performance does not reflect what can realistically be expected when military personnel assume the maintenance burden.

Table 3: Significant Problems and Limitations in Test Realism

	Army systems		Navy systems		Air Force systems	
Assessment questions	AHIP	Aquila	TLAM/C	DDG-51	IR Maverick	LANTIRN
Realism						
Operated by typical operational units?		X				X
Operated by typical operational personnel?	X	X	X		X	
Supported by typical support units?	X	X		X	b	X
Supported by typical support personnel?	X	X		X	b	X
Equipment put under realistic stress?	X	X	X	X	X	X
Personnel put under realistic stress?	X	X		X	X	X
Realistic combat tactics employed?	X	X	X	X	X	X
Physical environment approximates intended ranges?	X	X	X		X	X
Target systems approximate actual targets, realistically employed?	X			X	X	X
Threat systems approximate actual threat, realistically employed?	X	X	X	X	X	a
Tested system production representative and prepared for test in a realistic manner?				X		X

Note: empty cells signify "no significant problems or limitations found." X signifies "one or more significant problems or limitations found." a signifies "insufficient information to evaluate." b signifies "not applicable."

One case deserves special mention because it featured contractor involvement in operations as well as support. Two of the contractor's data collection technicians involved themselves in the conduct of the test on multiple occasions, despite warnings from officials. In at least one instance the contractor entered the crew area, unauthorized and unsolicited, and advised the crew while a mission was underway. Evidence of these actions, along with evidence of similar contractor involvement in maintenance, led the DOD Inspector General to confirm GAO's allegations of illegalities in the conduct of the OT. In the fiscal year 1987 authorization hearings, when the DOT&E director was presented with the case of an earlier test in which the system contractor actively participated in the test operations, he testified that DOT&E would ensure that this would not happen again. However, it did happen again in this instance, despite the fact that DOT&E personnel and their consultants conducted on-site monitoring of the test.

Directive 5000.3 states that testing should be conducted under conditions simulating combat stress. The DOT&E director also testified that adequacy of testing includes ensuring that tests are challenging. We recognize that there are safety and resource constraints that make some limitations inevitable and that most tests stressed equipment and personnel to at least some degree. Nevertheless, in all six cases, we found significant problems or limitations in the degree to which equipment was realistically stressed and, in five of six cases, in the degree to which

personnel were realistically stressed. Specific details are classified but, in general, instances of insufficient stress on equipment included: 1) the absence or significant underrepresentation of countermeasures (communication jamming, radar jamming, electro-optical countermeasures); 2) the use of tactics that facilitate performance during the test but are incompatible with system survivability in a realistic threat environment as defined by DOD; 3) use of targets that are hotter, slower, higher, more plentiful, less maneuverable, more likely to be stationary, or less likely to be camouflaged than DOD sources indicate would frequently be the case in combat, and 4) other instances where the outer edges of the specified performance envelope were not tested (although assets to test those edges existed). Instances of insufficient stress on personnel included crew familiarity with the test range or target area, assumptions that intelligence on enemy locations and other matters was readily available and accurate (in one case, needed meteorological data were obtained by a telephone call to the target area, an implausible method of data collection in wartime), various forms of cuing that reduced or eliminated the element of surprise, and failure to stress crew endurance commensurate with stated mission requirements. As a result, estimates of performance from the OT&Es tend to be biased upward, and performance under more realistic stress conditions remains unknown.

Finally, Directive 5000.3 also states that operational testing

should include threat-representative hostile forces. In testimony, the DOT&E director stated that adequacy of testing includes ensuring that the proper threat is being looked at; he stressed that the proper threat is the one that will be faced when the weapon system is fielded, not a "vintage-type" threat that is easily overcome. In at least some of the cases we reviewed, testers went to considerable lengths to portray the threat realistically, including validation and monitoring by service threat agencies. Still, in the five cases for which we had sufficient information to compare the threat portrayed in the test to the threat portrayed in DOD threat documents, all five showed significant problems or limitations. Again, the specific details are classified, but threat forces were in some cases less capable technologically than the Soviet forces that new U.S. systems would actually face, they were numerically underrepresented (that is, of low threat density), only partially portrayed (for example, ground threats present but air threats absent), otherwise not adequately depicted, or absent altogether for all or part of the test. As with stress and performance estimates, estimates of survivability from the OT&Es also tend to be biased upward, and survivability in a more realistic threat environment remains unknown.

Analysis and Service Test Agency Reporting

Findings on analysis and service test agency reporting are summarized in table 4. In five of six cases we found problems

with the various assumptions underlying the analyses of the test data. In at least two cases, such assumptions led to overly optimistic estimates of system capability, one of which was contradicted by available data. Other questionable analysis practices included combining data from disparate sources to yield overall performance estimates of unknown meaningfulness, removing valid data from performance computations, and lowering performance criteria after data collection. The impact of these practices was significant; they frequently allowed a system to appear to meet its performance requirements. Another problem was incomplete analyses--that is, analyses that did not integrate performance across all components of the total system or did not consider the limitations of other systems necessary for mission success.

Table 4:
Significant Problems and Limitations in Test Analysis and Reporting

	Army systems		Navy systems		Air Force systems	
Assessment Questions	AHIP	Aquila	TLAM/C	DDG-51	IR Maverick	LANTIRN
Analysis						
Measures quantitative and non-subjective?		X				X
Quantitative measures reliable and valid?	X	X	X			X
Analytic assumptions explicit and appropriate?	X	X		X	X	X
Sample size adequate to support statistically valid results?	X		X	X	X	X
Comparisions with other systems valid?	X	a	a	a	X	X
Service reporting						
Findings, conclusions, recommendations consistent with the evidence and appropriately qualified?	X	X	X	X	X	X
Reporting clear and comprehensive?			X	X	X	X

Note: empty cells signify "no significant problems or limitations found." X signifies "one or more significant problems or limitations found." a signifies "not applicable."

Within recent years, the Congress has indicated an interest in operational testing information that permits a comparison between the new system and the older system it is replacing. In three of the six cases, the system was tested comparatively against one or more older systems. However, comparisons were at times not tightly controlled, less challenging than comparisons against the new

system's own user criteria would have been, or without meaningful criteria altogether. In addition, some measures on which the older system would have performed better were either not included in the test or not assessed comparatively, and limitations of particular test scenarios or departures from realism were condoned on the assumption that they affected all systems equally, which was not always the case. In at least two cases, the limitation or departure clearly favored the new system.

There were also significant problems with each service test agency's reporting of results. In all six cases, the service test agencies stated findings, conclusions, or recommendations that were not consistent with the evidence or not sufficiently qualified. In four of these cases, one or more requirements were reported to have been met when they were not, and in one case the service test agency recommended full production despite numerous unresolved problems, one of which the service test agency itself had previously termed "urgent." One system was reported as showing "vast superiority" over its competitors in overall mission effectiveness when in fact it had demonstrated superiority in only two of the five mission areas being compared.

Conclusions

We found significant problems and limitations in the planning, execution, realism, analysis, and service test agency reporting of

the six OT&Es. Some of these problems and limitations were unavoidable due to time, resource, or safety constraints, although numerous others were not. We therefore conclude that for major, conventional systems that reached the B-LRIP milestone by the end of fiscal year 1987, the OT&E being conducted under DOT&E oversight was not methodologically adequate for assessing the operational effectiveness and suitability of weapon systems. OT&E has tended to yield more favorable assessments than are likely to be found when the weapons are employed in combat, which can lead to the funding of weapon systems whose operational effectiveness and suitability have not been demonstrated. In sum, OT&E under DOT&E oversight has fallen short of the objectives sought by the Congress when it established the office.

DOT&E IMPACT ON THE OT&E PROCESS

In order to better specify the unique contribution of DOT&E, we assessed the DOT&E impact on the OT&E process for the six cases. Our ability to evaluate DOT&E's impact on the OT&E of the six systems we reviewed was limited because much of the communication between DOT&E and other DOD components is informal and undocumented. As our March 1987 report noted, this lack of documentation makes it difficult to determine accurately how effectively DOT&E carries out some of its functions. In addition, we did not receive all the relevant documentation we requested. And since other sources gave us relevant DOT&E documents that DOT&E

did not identify or provide, there may well be additional documents that no source provided. Consequently, our evidence regarding DOT&E influence on the OT&E process is inconclusive.

Impact on the B-LRIP Milestone

At the B-LRIP milestone, DOT&E reports to the secretary of defense and the Congress on the adequacy of OT and the effectiveness and suitability of the system. Therefore, the B-LRIP milestone represents a major opportunity for DOT&E impact on the program.

In three of the six cases (TLAM/C, Aegis, and IR Maverick), we found no evidence that DOT&E influenced the B-LRIP milestone, other than to in effect support the production decision to the secretary of defense and the Congress. In the case of IR Maverick, several other DOD units raised significant concerns about the adequacy of testing and the operational effectiveness of Maverick. They presented these concerns to DOT&E before the B-LRIP decision meeting and at the meeting itself. DOT&E's B-LRIP report nevertheless stated that testing was adequate and effectiveness was satisfactory. In addition, we could find no evidence that DOT&E attempted either to defend its position or to respond to the concerns raised at the B-LRIP milestone meeting.

The fourth case (Aquila) was proposed for termination by the

Army before the B-LRIP milestone and without consultation with DOT&E.

In the fifth case (AHIP), DOT&E took the position at B-LRIP that, as tested, AHIP demonstrated an operationally effective capability in only one of the three roles planned for it. Based primarily on DOT&E's assessment, the decision was made to procure AHIP for that role only. This decision had meaningful consequences; it meant that 179 AHIPs would be procured rather than the 578 the Army had requested. However, three other DOD units were also critical of AHIP's performance; one had already recommended "only a conditional approval of limited production" based on the test results, and the other two told us that they would have objected had DOT&E assessed AHIP as effective in more than one role. Because four different offices delivered essentially the same message, the individual impact of DOT&E's position is unclear in this case.

In the sixth case (LANTIRN), we found no evidence of DOT&E impact at the B-LRIP milestone for the navigation pod other than to in effect support the production decision to the secretary of defense and the Congress. Concerning the targeting pod, DOT&E advised the Air Force that full production was not justified by the operational tests and that if the Air Force would defer full production, a B-LRIP report to the Congress would not be required. Instead, DOT&E would report to the Congress whenever the Air Force

proposed to exceed a rate of 81 pods per year. However, 81 pods was the Air Force's intended purchase for the first year of full production. Essentially, DOT&E offered the Air Force a choice between a negative B-LRIP report to the Congress and a redefinition of the B-LRIP rate to delay the report. The Air Force chose the latter and thus was able to adhere to its planned first-year, full-scale production schedule.

Conclusions

Due to the limitations stated earlier, our assessment of DOT&E impact is inconclusive.

In our six cases, we found no evidence of DOT&E impact in three major production decisions other than to in effect support the decision, no opportunity for impact in one, and in the other two, impact that was either indistinguishable from that of other DOD units or that was more apparent than real.

QUALITY OF DOT&E DISSEMINATION OF INFORMATION TO THE CONGRESS

The statute establishing DOT&E imposed two principal congressional reporting responsibilities: 1) a B-LRIP report stating whether OT&E was adequate and whether OT&E results confirm the item or components to be effective and suitable for combat, and 2) an annual report. The statute also requires the director to

respond to requests from the Congress for information regarding OT&E. In addition, DOT&E has initiated on its own the publication of monthly highlight reports that provide summary information on office activities and the progress of OT&E for specific programs. These are the sources used in our evaluation of DOT&E dissemination of information to the Congress.

Congressional concern about obtaining complete and accurate information on OT&E was a major reason for the provisions concerning dissemination of information to the Congress in the DOT&E legislation. To determine the completeness and accuracy of DOT&E statements to the Congress, we compared the facts as stated in DOT&E reports to those identified during our evaluation. Results are summarized in table 5.

Table 5:
Significant Problems in Completeness & Accuracy of DOT&E Reporting

	Army systems		Navy systems		Air Force systems	
DOT&E reporting	AHIP	Aquila	TLAM/C	DDG-51	IR Maverick	LANTIRN
OT&E adequacy						
Statements complete?	X	a	X	X	X	X
Statements accurate?	X	X		X	X	X
System effectiveness and suitability						
Statements complete?		a	X	X	X	X
Statements accurate?	X	a	X	X		X

Note: empty cells signify "no significant problems or limitations found." X signifies "one or more significant problems or limitations found." a signifies "insufficient information to evaluate."

DOT&E Statements on Adequacy of OT&E

We found one or more individual DOT&E statements on OT&E adequacy to be incomplete in five of six cases and inaccurate in five of six cases. By incomplete statement, we mean a statement that omitted information relevant to an assessment of adequacy.

Typically, such omissions consisted of the failure to report test limitations such as those discussed above. In some cases, the limitations identified by the service test agency were reported, but additional limitations were not (IR Maverick, TLAM/C); and in others, the limitations identified by the service test agency and additional limitations were not reported (AHIP, Aegis, and LANTIRN). Inaccurate statements included the following: tests were described as more challenging and realistic than they actually were (Aegis and LANTIRN), certain test assets were reported not to exist when in fact they did exist (IR Maverick), and the sufficiency of the test data was overstated (AHIP and Aquila).

We further assessed whether DOT&E's assessments of overall OT&E adequacy in the B-LRIP reports and other sources of information disseminated to the Congress were supported by the evidence. Of the six favorable adequacy assessments, we found that five were not supported by the evidence (AHIP, Aquila, TLAM/C, Aegis, and IR Maverick). In the sixth (LANTIRN navigation pod), we found no evidence inconsistent with DOT&E's assessment of overall adequacy. DOT&E made no overall adequacy statement for the LANTIRN targeting pod testing; however, the fiscal year 1987 DOT&E annual report made clear that before a favorable B-LRIP report can be written further tests are required. We concur with that assessment.

DOT&E Statements on System Effectiveness and Suitability

We found one or more individual DOT&E statements on system effectiveness and suitability to be incomplete in four of five cases and inaccurate in four of five cases. (There was no official statement on system effectiveness and suitability for Aquila.) Incomplete statements included the following: failure to mention as "urgent" a problem so characterized by the service test agency (TLAM/C), omitting key factors from an analysis, resulting in an unrealistically favorable performance assessment (IR Maverick), and omitting unfavorable test results (Aegis and LANTIRN). Inaccurate statements primarily consisted of overstatements of performance (AHIP, TLAM/C, Aegis, and LANTIRN); in each of these cases we found statements in which specific aspects of system performance were reported as more successful than the test results demonstrated. We found no statements that were inaccurate because they underrated performance.

We further assessed whether DOT&E's assessments of overall system effectiveness and suitability in the B-LRIP reports and other sources of information disseminated to Congress were supported by the evidence. Of the five favorable assessments of system performance, we found that four were not supported by the evidence (AHIP, TLAM/C, Aegis, and IR Maverick). In the fifth (LANTIRN navigation pod), we found no evidence inconsistent with DOT&E's assessment of system effectiveness and suitability. In the

case of the LANTIRN targeting pod, we concur with DOT&E's statement that more testing is needed to assess effectiveness and suitability.

Conclusions

Each of the official DOT&E reports to the Congress that we reviewed contained significant incomplete or inaccurate statements, and most contained both. In addition, the majority of favorable overall assessments of OT&E adequacy and of system effectiveness and suitability were not supported by the evidence. As noted earlier, some problems and limitations in operational testing are unavoidable; however, we know of no reason why those problems and limitations cannot be reported completely and accurately. The omissions, inaccuracies, and overall assessments consistently presented a more favorable presentation to the Congress of test adequacy and system performance than was warranted by the facts. We therefore conclude that for major, conventional systems that reached the B-LRIP milestone by the end of fiscal year 1987, the quality of DOT&E dissemination of information to the Congress has not provided the complete and accurate picture of weapon performance that the Congress needs to make weapon funding decisions. As such, it has fallen short of the objectives sought by the Congress when it established DOT&E.

Summary of Conclusions

In sum, for major, conventional systems that reached the B-LRIP milestone by the end of FY 1987, our conclusions are as follows:

1) The OT&E being conducted under DOT&E oversight has not been methodologically adequate for assessing the operational effectiveness and suitability of weapon systems, the fact that certain test limitations are unavoidable notwithstanding. Because of significant problems and limitations in the planning, execution, realism, analysis, and service test agency reporting of operational tests, OT&E has tended to yield more favorable assessments than are likely to be found when the weapons are employed in combat, potentially leading to the funding of weapon systems with undemonstrated operational effectiveness and suitability.

2) Regarding the specific impact of DOT&E on B-LRIP decisions, we found no DOT&E impact in three of six major production decisions other than its support of the decision, no opportunity for impact in one, and in the other two, impact that was either not distinguishable from that of other DOD units or that was more apparent than real. Our assessment of DOT&E impact on testing--specifically, how they influenced the testing process--was inconclusive.

3) The quality of DOT&E dissemination of information to the Congress has not provided the complete and accurate picture of weapon performance that the Congress needs to make weapon funding decisions. The omissions, inaccuracies, and overall assessments consistently resulted in a more favorable presentation to the Congress of test adequacy and system performance than was warranted by the facts. While certain problems and limitations in OT&E are unavoidable, we know of no reason why those problems and limitations cannot be reported completely and accurately.

Mr. Chairman, this concludes my prepared remarks. I would be happy to answer any questions that you or the other Members of the Panel might have.